UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/660,780	09/12/2003	Nambi Seshadri	58268.00224	5880

32294 7590 09/21/2007

SQUIRE, SANDERS & DEMPSEY L.L.P.
14TH FLOOR
8000 TOWERS CRESCENT
TYSONS CORNER, VA 22182

EXAMINER	
LERNER, MARTIN	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
09/21/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/660,780	SESHADRI, NAMBI	
	Examiner	Art Unit	
	Martin Lerner	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *27 July 2007*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1 to 21* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1 to 21* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *27 July 2007* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

2. Claims 1 to 3, 5 to 7, 9 to 11, and 13 to 15 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Morris* in view of *Verma et al. ('351)*.

Concerning independent claims 1, 5, and 9, *Morris* discloses a speech recognition method, device, and system, comprising:

"an audio signal receiver configured to receive audio signals from a speech source" – a user speaks to system 100, and system 100 captures the user's speech with speech input unit 104 (column 4, lines 15 to 19: Figures 1 and 2: Block 202); speech is an audio signal;

"a video signal receiver configured to receive video signals from the speech source" – a user speaks to system 100, and system 100 captures the user's image with video input unit 102 (column 4, lines 15 to 19: Figures 1 and 2: Block 202);

"a processing unit configured to process the audio signals and the video signals" – system 100 combines any captured speech or video and proceeds to process the

combined data stream in multi-sensor fusion/recognition unit 106 (column 4, lines 20 to 24: Figures 1 and 2: Block 204);

"a conversion unit configured to convert at least one of the audio signals and the video signals to recognizable information" – system 100 interprets any verbal input using the speech recognition functions of multi-sensor fusion/recognition unit 106; speech recognition is supplemented by visual information captured by video input unit 102, such as any interpreted facial expressions (e.g., lip-reading); a list of spoken words is generated from the verbal input (column 4, lines 25 to 31: Figures 1 and 2: Block 206); spoken words are recognizable information;

"an implementation unit configured to implement a task based on the recognizable information" – system 100 provides a response based upon whether the user has asked a question or made a statement; if a user has asked a question, then system 100 searches knowledge database 116 for a response to the objective question; a user may ask: "What is the weather in Phoenix, today?"; system 100 retrieves an answer, and the information is communicated as output via computer monitor and speakers (column 4, line 56 to column 5, line 24: Figure 3: Blocks 306, 308, 310, 312, 322); responding to a question by searching a knowledge database for a weather report for Phoenix, and outputting the weather report, is equivalent to implementing a task.

Concerning independent claims 1, 5, and 9, the only elements omitted by *Morris* are "detecting if the audio signal can be processed", processing the audio signals "if it is detected that the audio signals can be processed", and processing the video signals "if it is detected that at least a portion of the audio signal cannot be processed". *Morris*

discloses processing both the audio and video signals for multi-sensor fusion, so that better recognition can be obtained from speech input and video input. However, *Verma et al. ('351)* teaches a classifier for decision fusion, where inputs from audio and video are combined. Every incoming sample *j* of audio and video is associated with a sample confidence value L_{ij}, and a weight w_{ij} is assigned to each classifier as a function of an overall confidence. Where there are only two classifiers, one for audio and one for video, a linear summation of all weights is one. (Column 3, Lines 13 to 21: Column 4, Lines 25 to 44) Thus, as background noise degrades efficiency of information in an audio channel, a confidence weight w_{ij} for an audio channel goes to zero, and a confidence weight w_{ij} for a video channel goes to one. Measuring a confidence of an audio channel is equivalent to "detecting if the audio signals can be processed", and assigning a weight w_{ij} of zero to an audio signal when it is very noisy is equivalent to processing only the video signal "if it is detected that at least a portion of the audio signal cannot be processed". Alternatively, the audio signal is processed "if it is detected that the audio signals can be processed" because the weight w_{ij} assigned to an audio channel is not zero when channel noise is low. An objective is to improve a classification accuracy of a decision fusion application by assigning confidences, where a reliability of a classifier can vary from sample to sample. (Column 1, Lines 55 to 66) It would have been obvious to one having ordinary skill in the art to only process audio signals if it is detected that the audio signals can be processed as taught by *Verma et al. ('351)* in a multi-sensor fusion/recognition unit of *Morris* for a purpose of improving an accuracy of classification for a decision fusion application.

Concerning independent claims 13 to 15, similar considerations apply as to independent claims 1, 5, and 9. *Verma et al. ('351)* teaches assigning a sample confidence value L_{ij}, and a weight w_{ij} to each classifier, so that when an audio classifier determines that an audio channel is very noisy, a weight w_{ij} of zero is assigned to an audio channel, which corresponds to "detecting if the audio signals can be converted into a recognizable format". Implicitly, a noisy speech signal would not be recognizable by speech recognition, and, thus, could not be converted into a recognizable format.

Concerning claims 2, 6, and 10, *Morris* discloses that video input unit 102 receives face/voice expressions and interpreted facial expressions including lip-reading (column 4, lines 27 to 30: Figures 1 and 2).

Concerning claims 3, 7, and 11, *Morris* discloses that, in one embodiment, processing by multi-sensor fusion recognition unit 106 is split into three parallel processes to minimize time of processing (column 4, lines 20 to 24: Figures 1 and 2).

3. Claims 4, 8, and 12 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Morris* in view of *Verma et al. ('351)* as applied to claims 1, 5, and 9 above, and further in view of *Bakis et al.*

Morris does not expressly disclose a storage unit for storing the audio signals and the video signals to a destination source, and a transmitter for sending the audio signals and the video signals to a destination source. However, it is well known to

operate biometric identification via a client/server network, where biometric data is stored on a server, and biometric data is collected locally but compared to stored biometric data on the server. *Bakis et al.* teaches an analogous art method and apparatus for recognizing the identity of individuals by a speaker recognition system and a lip classifier, where biometric attributes are pre-stored for later retrieval so that they may be compared. Further, a server is included for interfacing with a plurality of biometric recognition systems to receive requests for biometric attributes therefrom and transmit biometric attributes thereto. The server has a memory device for storing the biometric attributes. (Column 8, Line 47 to Column 9, Line 16) Objectives are to provide a significant increase in the degree of accuracy of recognition and to provide a significant reduction in fraudulent or errant access to a service and/or facility. It would have been obvious to one having ordinary skill in the art to store and send biometric attributes to a server ("a destination source") as taught by *Bakis et al.* in a method, device, and system for combining audio and video signals of *Morris* for purposes of increasing accuracy of recognition and reducing fraudulent access.

4. Claims 16 to 18 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Morris* in view of *Verma et al. ('351)* as applied to claims 1, 5, and 9 above, and further in view of *Basu et al.*

Verma et al. ('351) discloses that an audio channel may be noisy, implying that an audio channel may be assigned a weight w_{ij} of zero, but omits "defining an error threshold", "comparing a number of errors detected in the audio signal with the

threshold", and "determining that the audio signals can not be processed if the number of detected errors equals or exceeds the threshold." However, thresholding is well known for a variety of purposes in speech processing. *Basu et al.* teaches a method and apparatus for audio-visual speech recognition, where a confidence estimation is performed to include a measurement of noise levels. A higher level of noise associated with a signal means that the confidence attributed to the recognition results associated with that signal is lower. Therefore, these confidence measures are taken into consideration during the weighting of the visual and acoustic results. (Column 8, Lines 25 to 41) Specifically, *Basu et al.* suggests that verification can be performed by score thresholding (column 6, lines 50 to 52), and errors in an audio decoding path may be detected, so that depending on the number of errors, a confidence measure may be produced (column 11, lines 28 to 31). Thus, *Basu et al.* suggests applying a threshold to a number of errors in an audio signal so as to assign a confidence to an audio channel. An objective is to provide further improvements for speaker recognition under acoustically degraded conditions including background noise. (Column 1, Lines 35 to 50) It would have been obvious to one having ordinary skill in the art to compare a number of errors in a speech signal to a threshold so as to apply a confidence measure as taught by *Basu et al.* in a method and apparatus for assigning a weight of an audio channel of *Verma et al. ('351)* for a purpose of improving speaker recognition under acoustically degraded conditions including background noise.

5. Claims 19 to 21 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Morris* in view of *Verma et al. ('351)* as applied to claims 1, 5, and 9 above, and further in view of *Brunelli et al.*

Verma et al. ('351) omits "determining if the video images of the user are detected", and "indicating to the user if the video image is not detected." However, one having ordinary skill in the art would understand that if the camera does not properly capture a face of a speaker in a method and apparatus for audio-visual speech recognition, then the camera would need to be adjusted. Specifically, *Brunelli et al.* teaches an integrated multisensory recognition system for speaker-recognition and visual-features recognition (Abstract), where an attention module 9 is sensitive to a signal provided by a television camera 3. When attention module 9 detects a face due to the arrival of a person P in front of television camera 3, a snapping module 10 waits until a scene in front of television camera 3 has stabilized, and checks that certain elementary condition are satisfied. When snapping module 10 has verified the existence of conditions of stability of a framed image, an acoustic indicator or loud speaker asks person P to utter certain words to initiate multisensory recognition. (Column 4, Line 50 to Column 5, Line 34: Figure 2) Thus, person P, or "the user", is notified when his/her images are not detected because an acoustic indicator does not prompt the user to speak the words; a user only hears an audio indication when his/her image is captured, so an absence of a prompt is equivalent to an indication that the video image was not detected. An objective is to combine acoustic and visual data in

an optimal manner that reduces probabilities of error to a minimum. (Column 2, Lines 3 to 10) It would have been obvious to one having ordinary skill in the art to provide a feature of notifying a user if a video image is not detected as taught by *Brunelli et al.* in a method and apparatus of audio and video decision fusion of *Verma et al. ('351)* for a purpose of combining acoustic and visual data in an optimal manner that reduces probabilities of error to a minimum.

Response to Arguments

6. Applicant's arguments filed 27 July 2007 have been considered but are moot in view of the new grounds of rejection, necessitated by amendment.

Conclusion

7. The prior art made of record and not relied upon is considered pertinent to Applicant's disclosure.

Boreczky et al., Yehia et al., Bangalore et al., Bellegarda et al., and Basson et al. disclose related art.

8. Applicant's amendment necessitated the new grounds of rejection presented in this Office Action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within

TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

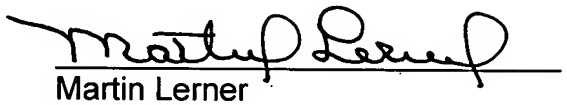
Any inquiry concerning this communication or earlier communications from the examiner should be directed to Martin Lerner whose telephone number is (571) 272-7608. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David R. Hudspeth can be reached on (571) 272-7843. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a

USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

ML
9/10/07



Martin Lerner
Examiner
Group Art Unit 2626